Part-II Parametric Signal Modeling and Linear Prediction Theory
2. Discrete Wiener Filtering

Electrical & Computer Engineering
University of Maryland, College Park

Acknowledgment: ENEE630 slides were based on class notes developed by Profs. K.J. Ray Liu and Min Wu. The LaTeX slides were made by Prof. Min Wu and Mr. Wei-Hong Chuang.
Contact: minwu@umd.edu. Updated: November 5, 2012.
Preliminaries

[ Readings: Haykin’s 4th Ed. Chapter 2, Hayes Chapter 7 ]

- Why prefer FIR filters over IIR?
  - FIR is inherently stable.

- Why consider complex signals?
  - Baseband representation is complex valued for narrow-band messages modulated at a carrier frequency. Corresponding filters are also in complex form.

\[ u[n] = u_I[n] + j u_Q[n] \]

- \( u_I[n] \): in-phase component
- \( u_Q[n] \): quadrature component

the two parts can be amplitude modulated by \( \cos 2\pi f_c t \) and \( \sin 2\pi f_c t \).
(1) General Problem

(Ref: Hayes §7.1)

Want to process $x[n]$ to minimize the difference between the estimate and the desired signal in some sense:

A major class of estimation (for simplicity & analytic tractability) is to use linear combinations of $x[n]$ (i.e. via linear filter).

When $x[n]$ and $d[n]$ are from two w.s.s. random processes, we often choose to minimize the mean-square error as the performance index.

$$\min_w J \triangleq \mathbb{E} [ |e[n]|^2 ] = \mathbb{E} \left[ |d[n] - \hat{d}[n]|^2 \right]$$
(2) Categories of Problems under the General Setup

1. Filtering
2. Smoothing
3. Prediction
4. Deconvolution
Wiener Problems: Filtering & Smoothing

- **Filtering**
  - The classic problem considered by Wiener
  - \( x[n] \) is a noisy version of \( d[n] \): \( x[n] = d[n] + v[n] \)
  - The goal is to estimate the true \( d[n] \) using a causal filter (i.e., from the current and post values of \( x[n] \))
  - The causal requirement allows for filtering on the fly

- **Smoothing**
  - Similar to the filtering problem, except the filter is allowed to be non-causal (i.e., all the \( x[n] \) data is available)
Wiener Problems: Prediction & Deconvolution

- **Prediction**
  - The causal filtering problem with $d[n] = x[n+1]$, i.e., the Wiener filter becomes a linear predictor to predict $x[n+1]$ in terms of the linear combination of the previous value $x[n], x[n-1], \ldots$

- **Deconvolution**
  - To estimate $d[n]$ from its filtered (and noisy) version $x[n] = d[n] \ast g[n] + v[n]$
  - If $g[n]$ is also unknown $\Rightarrow$ blind deconvolution. We may iteratively solve for both unknowns
FIR Wiener Filter for w.s.s. processes

Design an FIR Wiener filter for jointly w.s.s. processes \( \{x[n]\} \) and \( \{d[n]\} \):

\[
W(z) = \sum_{k=0}^{M-1} a_k z^{-k} \quad (\text{where } a_k \text{ can be complex valued})
\]

\[
\hat{d}[n] = \sum_{k=0}^{M-1} a_k x[n-k] = a^T x[n] \quad (\text{in vector form})
\]

\[
\Rightarrow e[n] = d[n] - \hat{d}[n] = d[n] - \sum_{k=0}^{M-1} a_k x[n-k]
\]

By summation-of-scalar:

\[
J = E[|e[n]|^2] = E[e[n]e^*[n]]
\]

\[
= E[|d[n]|^2] - E[d[n] \sum_{k=0}^{M-1} a_k^* x[n-k]] - E[d^*[n] \sum_{k=0}^{M-1} a_k x[n-k]] + E\left[\sum_{k=0}^{M-1} \sum_{l=0}^{M-1} a_k a^*_l E[x[n-k]x[n-l]]\right]
\]

\[
= E[|d[n]|^2] - \sum_{k=0}^{M-1} a_k^* E[d[n]x^*[n-k]] - \sum_{k=0}^{M-1} a_k E[d^*[n]x[n-k]] + \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} a_k a^*_l E[x[n-k]x[n-l]]
\]

\[
\hat{f}_X(l-k)
\]
FIR Wiener Filter for w.s.s. processes

In matrix-vector form:

\[ J = \mathbb{E} \left[ |d[n]|^2 \right] - a^H p^* - p^T a + a^H R a \]

where \( x[n] = \begin{bmatrix} x[n] \\
                            x[n-1] \\
                            \vdots \\
                            x[n-M+1] \end{bmatrix} \), \( p = \begin{bmatrix} \mathbb{E}[x[n]d^*[n]] \\
                                \vdots \\
                                \mathbb{E}[x[n-M+1]d^*[n]] \end{bmatrix} \), \( a = \begin{bmatrix} a_0 \\
                            \vdots \\
                            a_{M-1} \end{bmatrix} \).

- \( \mathbb{E} \left[ |d[n]|^2 \right] \): \( \sigma^2 \) for zero-mean random process
- \( a^H R a \): represent \( \mathbb{E} \left[ a^T x[n] x^H[n] a^* \right] = a^T R a^* \)
Perfect Square

1. If $R$ is positive definite, $R^{-1}$ exists and is positive definite.

2. $(Ra^* - p)^H R^{-1} (Ra^* - p) = (a^T R^H - p^H)(a^* - R^{-1}p)$
   
   
   $$= a^T R^H a^* - p^H a^* - a^T R^H R^{-1} p + p^H R^{-1} p$$
   
   Thus we can write $J(a)$ in the form of perfect square:

   $$J(a) = \mathbb{E} [ |d[n]|^2 ] - p^H R^{-1} p + (Ra^* - p)^H R^{-1} (Ra^* - p)$$
   
   Not a function of $a$; Represent $J_{\text{min}}$. >0 except being zero if $Ra^* - p = 0$
Perfect Square

\( J(a) \) represents the error performance surface:
convex and has unique minimum at \( R_a^* = p \)

Thus the necessary and sufficient condition for determining the optimal linear estimator (linear filter) that minimizes MSE is

\[ R_a^* - p = 0 \Rightarrow R_a^* = p \]

This equation is known as the **Normal Equation**.
A FIR filter with such coefficients is called a **FIR Wiener filter**.
Perfect Square

\[ \mathbf{R}_a^* = \mathbf{p} \quad \therefore \quad \mathbf{a}_{\text{opt}}^* = \mathbf{R}^{-1}\mathbf{p} \text{ if } \mathbf{R} \text{ is not singular (which often holds due to noise)} \]

When \{x[n]\} and \{d[n]\} are jointly w.s.s. (i.e., crosscorrelation depends only on time difference)

This is also known as the Wiener-Hopf equation (the discrete-time counterpart of the continuous Wiener-Hopf integral equations)
Principle of Orthogonality

Note: to minimize a real-valued func. $f(z, z^*)$ that’s analytic (differentiable everywhere) in $z$ and $z^*$, set the derivative of $f$ w.r.t. either $z$ or $z^*$ to zero.

- Necessary condition for minimum $J(a)$: (nece.& suff. for convex $J$)

$$\frac{\partial}{\partial a_k} J = 0 \text{ for } k = 0, 1, \ldots, M - 1.$$ 

$$\Rightarrow \frac{\partial}{\partial a_k} \mathbb{E} [e[n] e^*[n]] = \mathbb{E} \left[ e[n] \frac{\partial}{\partial a_k} (d^*[n] - \sum_{j=0}^{M-1} a_j x^*[n - j]) \right]$$

$$= \mathbb{E} \left[ e[n] \cdot (-x^*[n - k]) \right] = 0$$

**Principal of Orthogonality**

$$\mathbb{E} [e_{opt}[n] x^*[n - k]] = 0 \text{ for } k = 0, \ldots, M - 1.$$

The optimal error signal $e[n]$ and each of the $M$ samples of $x[n]$ that participated in the filtering are statistically uncorrelated (i.e., orthogonal in a statistical sense)
Principle of Orthogonality: Geometric View

Analogy:
r.v. $\Rightarrow$ vector;
$E(XY)$ $\Rightarrow$ inner product of vectors

$\Rightarrow$ The optimal $\hat{d}[n]$ is the projection of $d[n]$ onto the subspace spanned by \{x[n], \ldots, x[n - M + 1]\} in a statistical sense.

The vector form: 
\[ E[x[n]e^*_\text{opt}[n]] = 0. \]

This is true for any linear combination of $x[n]$ and for FIR & IIR:
\[ E[\hat{d}_\text{opt}[n]e_{\text{opt}}[n]] = 0. \]
Minimum Mean Square Error

Recall the perfect square form of $J$:

$$J(a) = \mathbb{E} \left[ |d[n]|^2 \right] - p^H R^{-1} p + (Ra^* - p)^H R^{-1} (Ra^* - p)$$

$$\therefore J_{\text{min}} = \sigma_d^2 - a_o^H p^* = \sigma_d^2 - p^H R^{-1} p$$

Also recall $d[n] = \hat{d}_{\text{opt}}[n] + e_{\text{opt}}[n]$. Since $\hat{d}_{\text{opt}}[n]$ and $e_{\text{opt}}[n]$ are uncorrelated by the principle of orthogonality, the variance is

$$\sigma_d^2 = \text{Var}(\hat{d}_{\text{opt}}[n]) + J_{\text{min}}$$

$$\therefore \text{Var}(\hat{d}_{\text{opt}}[n]) = p^H R^{-1} p$$

$$= a_o^H p^* = p^H a_o^* = p^T a_o \quad \text{real and scalar}$$
Example and Exercise

- What kind of process is $\{x[n]\}$?
- What is the correlation matrix of the channel output?
- What is the cross-correlation vector?
- $w_1 = ?$  $w_2 = ?$  $J_{\min} = ?$
Detailed Derivations
Another Perspective (in terms of the gradient)

Theorem: If \( f(z, z^*) \) is a real-valued function of complex vectors \( z \) and \( z^* \), then the vector pointing in the direction of the maximum rate of the change of \( f \) is \( \nabla_{z^*} f(z, z^*) \), which is a vector of the derivative of \( f() \) w.r.t. each entry in the vector \( z^* \).

Corollary: Stationary points of \( f(z, z^*) \) are the solutions to \( \nabla_{z^*} f(z, z^*) = 0 \).

Complex gradient of a complex function:

\[
\begin{array}{c|cc|c}
\nabla_z & a^H z & z^H a & z^H A z \\
\hline
\nabla_{z^*} & a^* & 0 & A^T z^* = (Az)^* \\
\end{array}
\]

Using the above table, we have \( \nabla_{a^*} J = -p^* + R^T a \).

For optimal solution: \( \nabla_{a^*} J = \frac{\partial}{\partial a^*} J = 0 \)

\( \Rightarrow R^T a = p^* \), or \( Ra^* = p \), the Normal Equation. \( \therefore \) \( a_{opt} = R^{-1} p \)

(Review on matrix & optimization: Hayes 2.3; Haykins(4th) Appendix A,B,C)
Review: differentiating complex functions and vectors

1. Differentiable at $z_0$

\[ \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} \text{ exist} \quad \Rightarrow \quad \text{Need to converge in all directions for } \Delta z \to 0 \]

Recall: $f(z)$ is analytic (i.e., differentiable everywhere) on region $D$ if $f(z) = u(x,y) + i v(x,y)$ is continuous and satisfy Cauchy-Riemann condition

\[ \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \& \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \]

2. e.g. $f(z) = |z|^2 = (x^2 + y^2) + i \times 0$

$\Rightarrow$ DOES NOT satisfy Cauchy-Riemann.
Review: differentiating complex functions and vectors

Unlike the real value, \( f(\bar{z}) = |z|^2 \) has unique minimum at \( z = 0 \), but not differentiable from complex analysis (any func. that depends on \( z^* \) is not differentiable).

Case, \( \frac{df(x)}{dx} = 0 \):

We can either minimize \( f(x, y) \) w.r.t. \( x \& y \) where \( z = x + iy \), or treat \( z \) and \( z^* \) as indep. variables and minimize \( f(z, z^*) \) w.r.t. both \( z \) and \( z^* \). i.e. \( \frac{\partial f}{\partial z} = 0 \) and \( \frac{\partial f}{\partial z^*} = 0 \).

Minimizing a real-valued func. of \( z \) and \( z^* \) (and the func. is analytic w.r.t. both \( z \) and \( z^* \)) is somewhat easier:

The optimal points may be found by setting the derivative of \( f(z, z^*) \) w.r.t. either \( z \) or \( z^* \) equal to zero and solve for \( z \).

E.g. \( f(z, z^*) = |z|^2 = z \cdot z^* \). Sufficient to have \( \frac{\partial f}{\partial z^*} = z = 0 \).
Differentiating complex functions: More details

\[ \delta = x + i y \]
\[ f(\delta) = u(x, y) + i v(x, y) \]

**Note:**
\[ x = \frac{1}{2} (\delta + \delta^*) \]
\[ y = \frac{1}{2i} (\delta - \delta^*) \]
\[ \Rightarrow \begin{cases} \frac{df}{dx} = \frac{2u}{\partial x} + i \frac{2v}{\partial x} \\ \frac{df}{dy} = \frac{2u}{\partial y} + i \frac{2v}{\partial y} \end{cases} \]

\[ \frac{df}{\delta} = \frac{1}{2} \left[ \frac{df}{dx} + \frac{df}{dy} (-i) \right] \quad \text{i.e.} \quad \frac{df}{\delta} = \frac{df}{dx} \frac{dx}{\delta} + \frac{df}{dy} \frac{dy}{\delta} \]

\[ \frac{df}{\delta^*} = \frac{1}{2} \left[ \frac{df}{dx} + \frac{df}{dy} i \right] \]

For real-valued \( f(\delta) \), i.e. \( f(\delta) = u(x, y) \), we have:
1. \( \frac{df}{dy} = \left( \frac{df}{dy^*} \right)^* \)
2. Gradient \( \nabla u = \begin{bmatrix} \frac{2u}{\partial x} \\ \frac{2u}{\partial y} \end{bmatrix} \Rightarrow \frac{\partial u}{\partial x} + i \frac{\partial u}{\partial y} = \frac{df}{dx} + i \frac{df}{dy} \)

**E.g. 1**
If \( f(\delta) = \delta = x + iy \)
\[ \frac{df}{\delta} = \frac{1}{2} \left[ \frac{df}{dx} + \frac{df}{dy} (-i) \right] = \frac{1}{2} (1 + i) = 0 ; \quad \frac{df}{\delta^*} = \frac{1}{2} \left( \frac{df}{dx} - \frac{df}{dy} i \right) = \frac{1}{2} (1 - i) = 1 \]

**E.g. 2**
If \( f(\delta) = |\delta|^2 \)
Let \( A \overset{df}{\text{def}} \frac{f(\delta + \Delta \delta) - f(\delta)}{\Delta \delta} = (\delta + \Delta \delta)(\delta^* + (\Delta \delta)^*) - \delta \cdot \delta^* = 8 + (\Delta \delta)^* + \delta (\Delta \delta)^* \)

For \( \Delta \delta = \Delta x + 0 \cdot i \):
\[ \lim_{\Delta x \to 0} \frac{8}{\Delta x} = 1 \quad \Rightarrow \quad A \to \delta^* \delta \]

For \( \Delta \delta = 0 + \Delta y \cdot i \):
\[ \lim_{\Delta y \to 0} \frac{8}{\Delta y} = -1 \quad \Rightarrow \quad A \to \delta^* \delta \]

\( \therefore \) \( A \) converges to different results for different directions as \( \Delta \delta \to 0 \) except for \( \delta = 0 \)

\( \therefore \) The limit doesn't exist, except for \( \delta = 0 \)

(and thus not differentiable)